Ownership Structure, Institutional Organization and Measured X-Efficiency

By KENNETH J. BUTTON AND THOMAS G. WEYMAN-JONES*

Efficiency measurement has become a very popular field in applied economics in recent years, and with this interest there has been a large intellectual investment in refining the empirical methods available to researchers in the area.[1] In this paper, we relate these developments to Harvey Leibenstein's original 1966 insight into the psychological ideas underlying the notion that economic agents may not achieve maximal efficiency in their productive decisions and behavior. Of course, it is always possible to argue that apparent inefficiency only arises from a failure of the observer to realize what it is that is being maximized. However, we shall evade this easy escape route into nonfalsifiable hypothesizing and instead shall take at face value the fact that too many empirical studies have come up with substantial measures of inefficiency for us to ignore its importance for normative economics.

I. X-Efficiency and Technical Efficiency

Despite this ongoing concern to explore the causes of inefficiency, there has sometimes been a tendency to use jargon rather too generally. People, for instance, sometimes have the tendency use the terms X-efficiency and technical efficiency interchangeably.[2] In this respect, in their work on X-efficiency and technical efficiency, Leibenstein (1966) and Michael J. Farrell (1957) were both seeking to explain why firms may not be minimizing their costs of production. While similar in their orientation, there are in fact important distinctions in the economic theories underlying X-efficiency and technical efficiency, as Leibenstein has himself pointed out on several occasions:

> The concept of T.E. [technical efficiency] suggests that the problem is a technical one and has to do with the techniques of an input called management. Under X.E. [X-efficiency] the basic problem is viewed as one that is *intrinsic* to the nature of human organization, both organization within the firm and organization outside of the firm. [italics as in original]
>
> [Leibenstein, 1977 p. 312][3]

> X-efficiency is not the same thing as what is frequently referred to as technical efficiency, since X-efficiency may arise for reasons outside the knowledge or capability of management attempting to do the managing .... In other words, it is not only a matter of techniques of management, or anything else "technical" in carrying out decisions, that is involved in X-efficiency. (Leibenstein, 1980 p. 27)

The key distinction is, therefore, in the objectives of their work. Farrell was essentially concerned with empirical matters and, in particular, with measurement (i.e., "to provide a satisfactory measure of productive

*Applied Microeconomics Research Group, Department of Economics, Loughborough University, Loughborough, LE11 3TU, United Kingdom.

[1]This is illustrated for example by the October 1990 special supplement to the *Journal of Econometrics* (A. Y. Lewin and C. A. K. Lovell, 1990), which is wholly devoted to the comparison of nonparametric and stochastic approaches.

[2]For example, see Kenneth H. Shapiro and Jürgen Müller (1977) on agricultural production in Tanzania.

[3]In this paper, he further elaborates on this point by highlighting seven important ways in which the concepts differ.

efficiency... and to show how it can be computed in practice" [Farrell, 1957 p. 11]). The finer points of motivation and managerial objectives were not part of Farrell's interests. In contrast, while Leibenstein in a succession of papers does offer some guidance to levels of X-inefficiency, albeit often almost anecdotal, his main concern was in establishing a better understanding of the way decisions are arrived at and, to this end, to examine relevant aspects of psychological and physiological aspects of human nature. In this sense, his work challenged and, at the same time, enlarged upon the basic assumptions of neoclassical economics (Mark Perlman, 1990). In contrast, Farrell was more interested in measuring certain observed phenomena, but within an essentially well-established neoclassical economic modeling framework.

It is sufficient for our purposes to accept that many writers have either identified the concepts of technical and X-efficiency with each other, or distinguished them solely on the basis of the degree to which the maximality principle has characterized the underlying behavioral assumptions. For the purposes of this paper, one of the questions we shall ask is to what extent measured technical efficiency provides some insight into the ideas of X-efficiency.

II. The Basic Propositions of X-Efficiency

In his 1978 *AER* paper on the basic proposition of X-efficiency theory, Leibenstein identified nonmaximizing behavior as the key to the idea of X-efficiency. This is a consequence of the nexus of pressures from the external environment on individual decision-makers and the responsibility consequences, or constraint concern, applying to the individual. The lower the intensity of environmental pressure on a decision-maker, the less is his or her concern with the constraints operating on the organization, and consequently, the lower is the effort expended. This reduced effort leads to higher costs and the basic proposition: the looser the effort-responsibility consequences, the greater the degree of X-inefficiency (the excess of actual over minimum cost, or the difference between maximal effectiveness of utilization and actual utilization).

Writers on X-efficiency theory, following Leibenstein's lead, have identified many possible sources for the failure of the environmental pressures on decision-makers to call forth maximal effort. These include the difficulties of principal–agent relationships in hierarchies within organizations. Indeed, Leibenstein seems to have been one of the first economists explicitly to mention principal–agent relationships as an important source of inefficiency. A classic example in the literature on principal–agent games arises in Jean-Jacques Laffont and Jean Tirole's (1986) model of adverse selection and moral hazard, one equilibrium of which has a regulator extracting allocatively efficient behavior from an asymmetrically well-informed monopolist with unknown costs and unobserved effort, but with costs set above and effort set below the first-best level. The rent from the firm's information monopoly accrues as pure X-inefficiency. Relationships between principal and agent often result in incomplete contingent contracts, and these can allow firms to evade the consequences of cost overruns, which are other manifestations of X-inefficiency. The entrepreneurship structure itself may be critical, with the classic issue of the separation of ownership from control being regarded as one of the earliest and most important sources of X-inefficiency.[4] Market-structure aspects are clearly critical in determining the extent to which constraints from the operating environment may bite on decision-makers. The degree of competitiveness in a firm's market, the extent to which it is incorporated as part of a public-sector bureaucracy, and the nature of the regulatory regime under which a firm operates are all primary sources of possible X-inefficiency.

[4] Leibenstein (1975) is the standard reference on ownership and control, but his more recent book (Leibenstein, 1987) provides a rather more thorough exposition of his views.

It should also be remembered that concern over potential X-inefficiency has become more than a point of theoretical discussion. The idea of X-efficiency has, for instance, had an enormous impact in several policy areas of which the design of the regulatory regime for the privatized public utilities (natural gas, telecommunications, electricity, and water supply) in Great Britain is one of the more important.[5] The old British nationalized industries were instructed to pursue allocative efficiency by pricing at marginal cost but were largely responsible for determining that marginal cost in what amounted to a formula of almost complete cost plus contract. This has been replaced by a price-cap mechanism (the so-called, "RPI-X formula") in which the utility is the residual claimant to the profits obtained by keeping costs below the cap (see Michael E. Beesley and Stephen C. Littlechild, 1989). The emphasis of regulation has moved from the confiscation of profits for allocative-efficiency reasons to maximizing the incentive to reduce costs. The immediate consequences have been substantial reductions in labor input used and, in the case of electricity, a switch from high-profile and complex capital-intensive technologies like nuclear power, which offer status and prestige to the industry's engineering management, to small-scale low-capital-cost technologies such as gas turbines. Of course, principal-agent and X-efficiency issues remain in the relationships between the utility managers and stockholders and between the regulators and the consumers.

III. Measured Efficiency

Although many different studies have been made of levels of X-efficiency, it is only in the last 12 years or so that really systematic efforts have been made to develop different empirical techniques for this purpose.[6] The econometric literature emphasizes three broad approaches which differ in their assumptions regarding the nature of the variation in a sample of firms or other decision-making organizations. All, however, either explicitly or implicitly draw on the measurement concepts introduced by Farrell, and all have been concerned with measuring the concept of an efficient frontier and the distance from it of the different organizations in the sample.[7]

It is our contention that, in concentrating on the issue of measurement, many of these econometric-efficiency studies have failed to bring out the importance of Leibenstein's ideas on the causes of underperformance and that this can largely be attributed to a failure of experimental design. However, we make an attempt to go behind the measurement calculations in the published studies to see whether it is possible to relate the results obtained to the factors likely to result in X-efficiency.

The three most popular techniques are parametric programming (Dennis J. Aigner and S. Chu, 1968), nonparametric programming (initiated by A. Charnes et al. [1978], who explicitly refer to X-efficiency), and parametric stochastic, or composed-error, frontiers (Aigner et al., 1977). The current state of the art is well reflected in the recent *Journal of Econometrics* supplement (Lewin and Lovell, 1990).

Among these three techniques, there are many possible categorizations and taxonomies, but a fundamental difference between the second technique and the other two relates to their assumptions regarding maximizing behavior. In both the first and

[5] The desire to reduce the perceived X-inefficiency associated with public-transport subsidies in Britain, and the subsequent initiation of individual bus-route tendering arrangements to replace block network subsidies in 1985 can be cited as another example (U.K. Department of Transport, 1984). The immediate impact was a significant reduction in the monies required to subsidize a given level of bus services.

[6] Roger S. Frantz (1988) surveys much of the empirical work explicitly on X-inefficiency, looking at the evidence in terms of the regulated industries, ownership form, and market structure.

[7] There also exists a small body of work which has been rather more concerned with trying to identify, by means of soft modeling, those industrial sectors that are most likely to be prone to X-inefficiency (e.g., Button, 1985).

third sets of techniques, the researcher postulates a parametric frontier based on a behavioral maximization hypothesis. This is usually a production frontier, a cost frontier, or a profit frontier. The production frontier assumes the existence of a functional relationship, $f(\mathbf{x})$, describing the maximum output obtainable from a vector of inputs, $\mathbf{x}$. The observed output of the typical firm falls short of this maximum output by an amount, ε equal to its technical inefficiency:

$$(1) \qquad \ln y = \ln f(\mathbf{x}) + \varepsilon \qquad \varepsilon \leq 0.$$

Duality relationships may be used to construct a parametric cost function for given input prices, $\mathbf{w}$, or the parametric profit function for given input and output prices, $\mathbf{w}$ and p. In either case, there is an implicit assumption that maximizing behavior is present and that it is exhibited by the most efficient firms in the sample. Inefficiency is measured by the size of the error between observed cost or profit performance and the parameterized maximum given by the functional relationship:

$$(2) \qquad \mathbf{w}'\mathbf{x} = C(y, \mathbf{w}) + \varepsilon \qquad \varepsilon \geq 0$$

or

$$(3) \quad py - \mathbf{w}'\mathbf{x} = \pi(\mathbf{w}, p) + \varepsilon \qquad \varepsilon \leq 0.$$

In the case of parametric stochastic frontier measurement, not all of the deviation of observed from maximal performance is attributed to inefficiency. Instead, a composed-error assumption is made that partitions ε between an asymmetrically distributed inefficiency term, u, and a symmetrically distributed noise term, v. Different estimation methods for such models are suggested, for example, by William H. Greene (1980) and Rodney Stevenson (1980).

What many of the empirical studies using these techniques fail to do is ask what a priori grounds exist for assuming that the best-practice firms in the sample (whose observations trace out the frontier) actually adopt optimizing behavior. It is almost universally true that researchers choose an industry-wide data set that is of intrinsic interest to them. They then investigate the degree of measured inefficiency against an optimized frontier without setting any controls on the nature of the constraint pressures that Leibenstein has argued will determine the degree of inefficient behavior. For example, in the composed-error cost frontier, measured inefficiency is:

$$(4) \quad u = \varepsilon - v$$
$$= (\text{observed cost} - \text{parameterized}$$
$$\text{frontier minimum cost}) - \text{noise}.$$

However in different industries and at different times, the parameterized frontier minimum cost may contain considerable unmeasured X-inefficiency, depending on the degree to which constraint concern pressures are operating. In this respect the nonparametric programming approach may have an additional attraction.[8] The nonparametric programming approach (also known as data-envelopment analysis [DEA]) proceeds by constructing the convex hull of the observed input–output observations for a given set of firms or organizations, under different assumptions about free disposability and returns to scale. For example, if $\mathbf{X}$ and $\mathbf{Y}$ are all the observations on inputs and outputs in an industry-wide sample of n firms, and $\mathbf{x}$ and y are the corresponding observations of a typical firm, then that firm's efficiency index, θ, assuming free disposability and variable returns to scale, is the solution to the linear program,

$$\text{choose } \{\theta, \boldsymbol{\lambda}\} \text{ to: } \min \theta \text{ such that:}$$
$$\theta \mathbf{x} \geq \boldsymbol{\lambda}'\mathbf{X}$$
$$y \leq \boldsymbol{\lambda}'\mathbf{Y}$$
$$\lambda_i \geq 0 \quad \sum \lambda_i = 1 \quad i = 1, \ldots, n.$$

[8]There is already a well-known set of arguments about the trade-off between imposing a parametric structure and assuming that all variation is due to inefficiency.

As such this technique constructs a frontier based simply on the distance of the best-practice firms from the rest. There is no implicit assumption of maximizing behavior on the part of any of the firms, including the best-practice firms on the frontier. It is still the case, though, that many DEA or nonparametric efficiency studies ignore the nature of the efficiency pressures likely to be operating on any given sample.

IV. Measured X-Efficiency and Efficiency Pressure

As we argued above, a failure of experimental design causes many researchers to choose their samples before or without considering whether efficiency incentives are likely to be operating. To determine the extent to which this might be important, we surveyed a selection of efficiency studies using all three of the broad techniques (described in more detail in Button and Weyman-Jones [1992]).

In all instances, the degree of measured inefficiency is very sensitive to the researcher's assumptions about the appropriate method of analysis. For example, in many cases the stochastic-frontier approach and the nonparametric-programming approach not only yield very different estimates of inefficiency for the same sample, but the distribution of inefficiency itself varies according to the measurement method used. One particularly nonrobust area, as Stevenson points out, is the choice of density function and its truncation point for the one-sided error in the stochastic-frontier model. Very severe assumptions are often made, including the restriction that the mass of the inefficiency density is most concentrated at zero. This sensitivity to assumptions makes comparisons of the results from different studies problematic, and to proceed further, so as to relate the empirical results to Leibenstein's work, requires us to be very selective in the choice of samples to compare.

For the purposes of this paper, we concentrate on a sample of published nonparametric studies, on the grounds that they do not make an obvious maximizing assumption in their method. We have no option but to accept at face value the researcher's own description of their samples and their reasons for their choice of sample, but it is clear that the majority were preoccupied with "measurement" (as in the Farrell approach), rather than "explanation" (in the Leibenstein tradition).

Using L. M. Seiford's (1990) bibliography, we examined a number of DEA efficiency studies, selecting those for which researchers clearly described the nature of the industry sampled and reported individual efficiencies or their sample statistics in the form of mean, standard deviation, and minimum.

An immediate problem arises in the choice of sample size used by researchers. It is well known that measured DEA efficiency in small samples is sensitive to the difference between the number of firms and the sum of inputs and outputs used. This is because the small number of free dimensions remaining increases the chance of each firm being seen as efficient. We therefore disregarded studies with sample sizes corresponding to less than about 35 degrees of freedom in the sense just described. This left a surprisingly small number of studies with usable, or comparable, results.

To relate these results to Leibenstein's work, we tried to determine for each study the researchers' own description of the nature of constraint-concern pressures operating on the industry at the time of the efficiency measurement. Our reduced sample included studies of U.S. and European financial institutions and government and other public services. We chose to construct a single indicator of the constraint-concern pressures, Z, which took a value of 0 if the industry was described by its investigators as competitive, privately owned, and not severely regulated (high constraint concern) and took a value of 1 if the industry was publicly owned or bureaucratically organized (low constraint concern). Surprisingly few research studies bother to describe such simple broad characteristics for their sample, which is particularly curious considering that efficiency is the focus of concern.

TABLE 1—SUMMARY STATISTICS FOR NINE LARGE SAMPLE EFFICIENCY STUDIES

Study	Efficiency			Bureaucracy (index of constraint concern)
	Mean	SD	Minimum	
1	0.650	0.180	0.180	0
2	0.790	—	—	0
3	0.700	0.050	0.600	0
4	0.896	0.080	0.628	0
5	0.771	0.130	0.408	0
6	0.710	0.180	0.329	1
7	0.906	0.144	0.618	1
8	0.609	0.149	0.175	1
9	0.973	0.046	0.840	0

TABLE 2—RANK CORRELATIONS BETWEEN EFFICIENCY MEASURES AND THE INDEX OF LACK OF CONSTRAINT CONCERN, Z

Mean efficiency	Standard deviation	Minimum efficiency
−0.18	0.57	−0.39

Fortunately, in our sample of samples, the distinction between the binary values seemed rather obvious.

If the ideas of X-efficiency are prevalent, then we should expect to find that where constrain concern was attenuated, mean efficiency is low, minimum efficiency is low, and the spread of efficiency among the survivor organizations or firms is high. This is broadly our finding, though the results can at best be described as suggestive. Table 1 sets out the surveyed results, and Table 2 reports the rank correlations with the index of constraint concern.

Our broad conclusion is that there is some suggestion that bureaucratic or publicly administered industries are on average less efficient, have lower extremes of efficiency, and show a wider dispersion of efficiency than privately owned, competitive, or weakly regulated industries. Put another way, the causes of inefficiency are out there waiting *to be measured*, but the ad hoc, or arbitrary, selection of samples that is characteristic of many efficiency studies makes it difficult to go beyond mere measurement into the nature of cause and effect.

V. Experimental Design

The efficiency literature contains two broad themes. On the one hand, there is a focus on measurement, in which some enormous advances in technique have been made recently. On the other hand, there is the explanation of cause and effect for which Harvey Leibenstein was the pioneering spirit. What is missing, at present, is a serious effort by investigators of efficiency measurement to relate their choice of sample or experimental design to tackling the issues raised by Leibenstein. We have measurement, and we have theory; but at present the two are not being related systematically. We have discovered weak but suggestive evidence that this would be a fruitful exercise. Perhaps this will be a major challenge for the next 25 years of X-efficiency theory.

REFERENCES

Aigner, Dennis J. and Chu, S., "On Estimating the Industry Production Function," *American Economic Review*, September 1968, *58*, 826–9.

______, **Lovell, C. A. Knox and Schmidt, Peter,** "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, July 1977, *6*, 21–37.

Beesley, Michael E. and Littlechild, Stephen C., "The Regulation of Privatized Monopolies in the U.K.," *Rand Journal of Economics*, Autumn 1989, *20*, 454–72.

Button, Kenneth J., "Potential Differences in the Degree of X-Inefficiency between Industrial Sectors in the United Kingdom," *Quarterly Review of Economics and Business*, Autumn 1985, *25*, 85–95.

______ **and Weyman-Jones, Thomas G.,** "X-Efficiency and Technical Efficiency," Economic Research Paper 92/1, Loughborough University, 1992.

Charnes, A., Cooper, W. W. and Rhodes, E., "Measuring the Efficiency of Decision-making Units," *European Journal of Operations Research*, November 1978, *2*, 429–44.

Farrell, Michael J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, 1957, *120* (3), 253–82.

Frantz, Roger S., *X-Efficiency: Theory, Evidence and Applications*, Boston: Kluwer, 1988.

Greene, William H., "On the Estimation of a Flexible Frontier Production Model," *Journal of Econometrics*, May 1980, *13*, 101–15.

Laffont, Jean-Jacques and Tirole, Jean, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, June 1986, *94*, 614–41.

Leibenstein, Harvey, "Allocative Efficiency vs. 'X-Efficiency'," *American Economic Review*, June 1966, *56*, 392–415.

______, "Aspects of the X-Efficiency Theory of the Firm," *Bell Journal of Economics*, Autumn 1975, *6*, 580–606.

______, "X-Efficiency, Technical Efficiency, and Incomplete Information Use: A Comment," *Economic Development and Cultural Change*, January 1977, *25*, 311–16.

______, "On the Basic Propositions of X-Efficiency Theory," *American Economic Review*, May 1978 (*Papers and Proceedings*), *68*, 328–32.

______, *Inflation, Income Distribution and X-Efficiency Theory*, London: Croom Helm, 1980.

______, *Inside the Firm—The Inefficiencies of Hierarchy*, Cambridge, MA: Harvard University Press, 1987.

Lewin, A. Y. and Lovell, C. A. K., eds., *Frontier Analysis: Parametric and Nonparametric Approaches, Journal of Econometrics*, October-November 1990, *46*, 1–245.

Perlman, Mark, "The Evolution of Leibenstein's X-Efficiency Theory," in Klaus Weiermair and Mark Perlman, eds., *Studies in Economic Rationality — X-Efficiency Examined and Extolled*, Ann Arbor: University of Michigan Press, 1990, pp. 7–26.

Seiford, L. M., "A Bibliography of Data Envelope Analysis (1978–90): Version 5.0," technical report, Department of Industrial Engineering, University of Massachusetts, Amherst, 1990.

Shapiro, Kenneth H. and Müller, Jürgen, "Sources of Technical Efficiency: The Roles of Modernization and Information," *Economic Development and Cultural Change*, January 1977, *25*, 293–310.

Stevenson, Rodney, "Likelihood Functions for Generalised Stochastic Frontier Estimation," *Journal of Econometrics*, May 1980, *13*, 57–66.

U.K. Department of Transport, *Buses*, Cmnd 9300, London: Her Majesty's Stationers Office, 1984.